Spectral Methods in Gaussian Modelling

Topic 2: Kernel Design

James Requiema and Wessel Bruinsma

University of Cambridge and Invenia Labs

21 January 2019
RFFs alleviate the $O(N^3)$ scaling.
+ RFFs alleviate the $O(N^3)$ scaling.

− RFFs do not help with choice of kernel.
RFFs alleviate the $O(N^3)$ scaling.

- RFFs do not help with choice of kernel.

How to parametrise a flexible kernel?
• Bochner’s Theorem:

\[ k(\tau) \xleftrightarrow{\mathcal{F}} s(\omega) = \text{PSD}. \]
Kernel Design (2)

- Bochner’s Theorem:

\[ k(\tau) \xrightarrow{\mathcal{F}} s(\omega) = \text{PSD}. \]

- PSD:
  - distribution of power contained in frequencies,
  - must be nonnegative and symmetric.
Kernel Design (2)

• Bochner’s Theorem:

\[ k(\tau) \xleftarrow{\mathcal{F}} s(\omega) = \text{PSD}. \]

• PSD:
  - distribution of power contained in frequencies,
  - must be nonnegative and symmetric.

• Easier to flexibly parametrise PSD!
• SSA (Lázaro-Gredilla et al., 2010) models PSD with symmetric average of lines:

\[ s(\omega) = \frac{1}{2Q} \sum_{q=1}^{Q} (\delta(\omega - \mu^{(q)}) + \delta(\omega + \mu^{(q)})). \]
• SSA (Lázaro-Gredilla et al., 2010) models PSD with symmetric average of lines:

\[ s(\omega) = \frac{1}{2Q} \sum_{q=1}^{Q} (\delta(\omega - \mu^{(q)}) + \delta(\omega + \mu^{(q)})). \]

• Inverse Fourier transform gives kernel:

\[ k(\tau) = \frac{1}{Q} \sum_{q=1}^{Q} \cos(\mu^{(q)^T} \tau). \]
• SSA (Lázaro-Gredilla et al., 2010) models PSD with symmetric average of lines:

\[ s(\omega) = \frac{1}{2Q} \sum_{q=1}^{Q} (\delta(\omega - \mu^{(q)}) + \delta(\omega + \mu^{(q)})) . \]

• Inverse Fourier transform gives kernel:

\[ k(\tau) = \frac{1}{Q} \sum_{q=1}^{Q} \cos(\mu^{(q)^T} \tau) . \]

• Strong parametric assumption: \( f(t) = \) sum of sines.
• SMK (Wilson and Adams, 2013) models PSD with symmetric mixture of Gaussians:

\[
s(\omega) = \frac{1}{2} \sum_{q=1}^{Q} w^{(q)} \left( \mathcal{N}(\omega; \mu^{(q)}, \Sigma^{(q)}) + \mathcal{N}(\omega; -\mu^{(q)}, \Sigma^{(q)}) \right).
\]
• SMK (Wilson and Adams, 2013) models PSD with symmetric mixture of Gaussians:

\[
s(\omega) = \frac{1}{2} \sum_{q=1}^{Q} w(q) \left( \mathcal{N}(\omega; \mu(q), \Sigma(q)) + \mathcal{N}(\omega; -\mu(q), \Sigma(q)) \right).
\]

• \(w(q) = 1/Q\) and \(\Sigma(q) \to 0\) recovers SSA.
SMK (Wilson and Adams, 2013) models PSD with symmetric mixture of Gaussians:

\[ s(\omega) = \frac{1}{2} \sum_{q=1}^{Q} w(q) \left( \mathcal{N}(\omega; \mu(q), \Sigma(q)) + \mathcal{N}(\omega; -\mu(q), \Sigma(q)) \right). \]

- \( w(q) = 1/Q \) and \( \Sigma(q) \to 0 \) recovers SSA.

- Inverse Fourier transform gives kernel:

\[ k^{(SMK)}(\tau) = \sum_{q=1}^{Q} w(q) \exp\left(-\frac{1}{2} \tau^T \Sigma(q) \tau\right) \cos\left(\mu(q)^T \tau\right), \]
• Equivalent generative model as a truncated Fourier series:

\[ f^{(SMK)}(t) = \sum_{q=1}^{Q} \sqrt{\omega(q)} (c_1^{(q)}(t) \cos(\mu^{(q)T} t) + c_2^{(q)}(t) \sin(\mu^{(q)T} t)), \]

\[ c_1^{(q)}, c_2^{(q)} \sim \mathcal{GP}(0, \exp(-\frac{1}{2} \tau^T \Sigma^{(q)} \tau)). \]
• Equivalent generative model as a truncated Fourier series:

\[
\begin{align*}
    f^{(\text{SMK})}(t) &= \sum_{q=1}^{Q} \sqrt{w(q)} (c_{1}^{(q)}(t) \cos(\mu^{(q)^{T}}t) + c_{2}^{(q)}(t) \sin(\mu^{(q)^{T}}t)), \\
    c_{1}^{(q)}, c_{2}^{(q)} &\sim \mathcal{GP}(0, \exp(-\frac{1}{2}\tau^{T}\Sigma^{(q)}\tau)).
\end{align*}
\]

• In SSA, \((c_{1}^{(q)}, c_{2}^{(q)})_{q=1}^{Q}\) are constant.
• Equivalent generative model as a truncated Fourier series:

\[ f^{(SMK)}(t) = \sum_{q=1}^{Q} \sqrt{w(q)} (c_1^{(q)}(t) \cos(\mu^{(q)^T}t) + c_2^{(q)}(t) \sin(\mu^{(q)^T}t)), \]

\[ c_1^{(q)}, c_2^{(q)} \sim \mathcal{GP}(0, \exp(-\frac{1}{2} \tau^T \Sigma(q) \tau)). \]

• In SSA, \((c_1^{(q)}, c_2^{(q)})_{q=1}^{Q}\) are constant.

• SMK fattens spectral lines by allowing \(c_1^{(q)}\) and \(c_2^{(q)}\) to vary with time.
Spectral Mixture Kernel (4)

+ Flexible, drop-in replacement
Spectral Mixture Kernel (4)

+ Flexible, drop-in replacement
+ Can recover many standard kernels
Spectral Mixture Kernel (4)

+ Flexible, drop-in replacement
+ Can recover many standard kernels
+ Models negative covariances
Spectral Mixture Kernel (4)

+ Flexible, drop-in replacement
+ Can recover many standard kernels
+ Models negative covariances
  - Unclear how many components needed
Spectral Mixture Kernel (4)

+ Flexible, drop-in replacement
+ Can recover many standard kernels
+ Models negative covariances

- Unclear how many components needed
- Hyperparameters difficult to optimise
• MOSMK (Parra and Tobar, 2017) generalises SMK to multiple outputs.
• MOSMK (Parra and Tobar, 2017) generalises SMK to multiple outputs.

• Uses multivariate extension of Bochner’s Theorem: Cramér’s Theorem.
MOSMK (Parra and Tobar, 2017) generalises SMK to multiple outputs.

Uses multivariate extension of Bochner’s Theorem: Cramér’s Theorem.

Multivariate PSD $S : \mathbb{R}^D \rightarrow \mathbb{C}^{P \times P}$. 
Multi-Output Spectral Mixture Kernel

- MOSMK (Parra and Tobar, 2017) generalises SMK to multiple outputs.
- Uses multivariate extension of Bochner’s Theorem: Cramér’s Theorem.
- Multivariate PSD $S: \mathbb{R}^D \rightarrow \mathbb{C}^{P \times P}$.
  - Must be symmetric: $S(\omega) = S^\dagger(-\omega)$, $S_{ii}(\omega) = S_{ii}(-\omega)$. 
• MOSMK (Parra and Tobar, 2017) generalises SMK to multiple outputs.

• Uses multivariate extension of Bochner’s Theorem: Cramér’s Theorem.

• Multivariate PSD $S: \mathbb{R}^D \rightarrow \mathbb{C}^{P \times P}$.
  
  • Must be symmetric: $S(\omega) = S^\dagger(-\omega)$, $S_{ii}(\omega) = S_{ii}(-\omega)$.
  
  • Must be nonnegative: $S(\omega) \geq 0$. 
• MOSMK models PSD with symmetric mixture of outer products of vectors of Gaussians:

\[
S(\omega) = \frac{1}{2} \sum_{q=1}^{Q} \left( R^{(q)}(\omega)R^{(q)\dagger}(\omega) + R^{(q)}(-\omega)R^{(q)\dagger}(-\omega) \right),
\]

\[
R_{i}^{(q)}(\omega) = w^{(q)} \exp \left( -\frac{1}{4}(\omega - \mu_{i}^{(q)})\Sigma_{i}^{(q)-1}(\omega - \mu_{i}^{(q)}) \right.
\]

\[
- \nu(\theta_{i}^{(q)T}\omega + \phi_{i}^{(q)}).\]

Inverse Fourier transform gives kernel:

\[
K_{ij}^{(MOSMK)}(\tau) = \sum_{q=1}^{Q} \alpha_{ij}^{(q)} \exp\left(-\frac{1}{2} (\tau + \theta_{ij}^{(q)})^T \Sigma_{ij}^{(q)} (\tau + \theta_{ij}^{(q)})\right) \\
\times \cos\left((\tau + \theta_{ij}^{(q)})^T \mu_{ij}^{(q)} + \phi_{ij}^{(q)}\right).
\]
• Equivalent generative model as truncated Fourier series:

\[ f_{i}^{(MOSMK)}(t) = \sum_{q=1}^{Q} w_{i}^{(q)} \left( c_{i1}^{(q)} (t - \theta_{i}^{(q)}) \cos(\mu_{i}^{(q)} T (t - \theta_{i}^{(q)}) + \phi_{i}^{(q)}) + c_{i2}^{(q)} (t - \theta_{i}^{(q)}) \sin(\mu_{i}^{(q)} T (t - \theta_{i}^{(q)}) + \phi_{i}^{(q)}) \right), \]

\[ \mathbb{E}[c_{ik}^{(p)}(t)c_{jl}^{(q)}(t')] = \begin{cases} 
\frac{\alpha_{ij}^{(q)}}{w_{i}^{(q)} w_{j}^{(q)}} \exp \left( -\frac{1}{2} (t - t')^{T} \Sigma_{ij}^{(q)} (t - t') \right) & \text{if } k = l, \ p = q, \\
0 & \text{otherwise.}
\end{cases} \]
• GSMK (Chen et al., 2018) generalises SMK to nonstationary signals.
• GSMK (Chen et al., 2018) generalises SMK to nonstationary signals.

• Uses the Gibbs kernel (Gibbs, 1997):

\[
k^{(\text{Gibbs})}(t, t') = \prod_{d=1}^{D} \sqrt{\frac{2\ell_d(t)\ell_d(t')}{\ell_d^2(t) + \ell_d^2(t')}} \exp \left( - \sum_{d=1}^{D} \frac{(t_d - t'_d)^2}{\ell_d^2(t) + \ell_d^2(t')} \right).
\]
Nonstationary EQ Kernel

- Cannot simply make length scale input dependent.
Nonstationary EQ Kernel

- Cannot simply make length scale input dependent.

- Construction of EQ from basis functions:

\[
\phi(t; c) = \left( \sqrt{\frac{2}{\pi \ell}} \right)^{\frac{1}{2}} \exp\left( -\frac{1}{\ell^2} (t - c)^2 \right),
\]
Generalised Spectral Mixture Kernel (2)

Nonstationary EQ Kernel

- Cannot simply make length scale input dependent.

- Construction of EQ from basis functions:

\[ \phi(t; c) = \left( \sqrt{\frac{2}{\pi \ell}} \right)^{\frac{1}{2}} \exp\left( -\frac{1}{\ell^2} (t - c)^2 \right), \]

\[ f(t) \mid n = \int_{-\infty}^{\infty} \phi(t; c)n(c) \, dc, \quad n(t) \sim \mathcal{GP}(0, \delta(t - t')), \]
Nonstationary EQ Kernel

- Cannot simply make length scale input dependent.

- Construction of EQ from basis functions:

\[
\phi(t; c) = \left(\sqrt{\frac{2}{\pi \ell}}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{\ell^2}(t - c)^2\right),
\]

\[
f(t) | n = \int_{-\infty}^{\infty} \phi(t; c)n(c) \, dc, \quad n(t) \sim \mathcal{GP}(0, \delta(t - t')),
\]

\[
\mathbb{E}[f(t)f(t')] = \int_{-\infty}^{\infty} \phi(t; c)\phi(t'; c) \, dc
\]
Nonstationary EQ Kernel

- Cannot simply make length scale input dependent.

- Construction of EQ from basis functions:

\[
\phi(t; c) = \left(\sqrt{\frac{2}{\pi \ell}}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{\ell^2}(t - c)^2\right),
\]

\[
f(t) \mid n = \int_{-\infty}^{\infty} \phi(t; c)n(c) \, dc, \quad n(t) \sim \mathcal{GP}(0, \delta(t - t')),
\]

\[
\mathbb{E}[f(t)f(t')] = \int_{-\infty}^{\infty} \phi(t; c)\phi(t'; c) \, dc
\]

\[
= \exp\left(-\frac{1}{2\ell^2}(t - t')^2\right).
\]
Nonstationary EQ Kernel

- Make length scale of $\phi$ dependent on $t$:

$$\phi(t; c) = \left( \sqrt{\frac{2}{\pi}} \frac{1}{\ell(t)} \right)^{\frac{1}{2}} \exp\left( -\frac{1}{\ell^2(t)} (t - c)^2 \right),$$
Nonstationary EQ Kernel

- Make length scale of $\phi$ dependent on $t$:

$$
\phi(t; c) = \left( \sqrt{\frac{2}{\pi}} \frac{1}{\ell(t)} \right)^{\frac{1}{2}} \exp \left( -\frac{1}{\ell^2(t)}(t - c)^2 \right),
$$

$$
f(t) \mid n = \int_{-\infty}^{\infty} \phi(t; c) n(c) \, dc, \quad n(t) \sim \mathcal{GP}(0, \delta(t - t')),
$$
Nonstationary EQ Kernel

- Make length scale of $\phi$ dependent on $t$:

$$
\phi(t; c) = \left(\sqrt{\frac{2}{\pi}} \frac{1}{\ell(t)}\right)^{\frac{1}{2}} \exp\left(- \frac{1}{\ell^2(t)} (t - c)^2\right),
$$

$$
f(t) \mid n = \int_{-\infty}^{\infty} \phi(t; c) n(c) \, dc, \quad n(t) \sim \mathcal{GP}(0, \delta(t - t')), 
$$

$$
\mathbb{E}[f(t)f(t')] = \int_{-\infty}^{\infty} \phi(t; c) \phi(t'; c) \, dc
$$
Generalised Spectral Mixture Kernel (3)

Nonstationary EQ Kernel

- Make length scale of $\phi$ dependent on $t$:

$$
\phi(t; c) = \left( \sqrt{\frac{2}{\pi}} \frac{1}{\ell(t)} \right)^{\frac{1}{2}} \exp \left( -\frac{1}{\ell^2(t)} (t - c)^2 \right),
$$

$$
f(t) | n = \int_{-\infty}^{\infty} \phi(t; c)n(c) \, dc, \quad n(t) \sim \mathcal{GP}(0, \delta(t-t')),
$$

$$
\mathbb{E}[f(t)f(t')] = \int_{-\infty}^{\infty} \phi(t; c)\phi(t'; c) \, dc
$$

$$
= \sqrt{\frac{2\ell(t)\ell(t')}{\ell^2(t) + \ell^2(t')}} \exp \left( -\frac{(t-t')^2}{\ell^2(t) + \ell^2(t')} \right).
$$
• GSMK replaces the EQs with Gibbs kernels:

\[ k^{(GSMK)}(t, t') = \sum_{q=1}^{Q} w^{(q)}(t) w^{(q)}(t') k_{q}^{(Gibbs)}(t, t') \times \cos\left(\mu^{(q)T}(t)t - \mu^{(q)T}(t')t'\right). \]
• GSMK replaces the EQs with Gibbs kernels:

\[
k^{(\text{GSMK})}(t, t') = \sum_{q=1}^{Q} w^{(q)}(t)w^{(q)}(t')k_{q}^{(\text{Gibbs})}(t, t') \\
\times \cos\left(\mu^{(q)\top}(t)t - \mu^{(q)\top}(t')t'\right).
\]

• \((w^{(q)}, \ell^{(q)}\mu^{(q)})_{q=1}^{Q}\) given log-GP priors.
• GSMK replaces the EQs with Gibbs kernels:

\[
k^{(\text{GSMK})}(t, t') = \sum_{q=1}^{Q} w^{(q)}(t)w^{(q)}(t')k_{q}^{(\text{Gibbs})}(t, t') \times \cos \left( \mu^{(q)T}(t)t - \mu^{(q)T}(t')t' \right).
\]

• \((w^{(q)}, \ell^{(q)}\mu^{(q)})_{q=1}^{Q}\) given log-GP priors.

• Estimated using MAP.
• Equivalent generative model as truncated Fourier series:

\[
 f^{(GSMK)}(t) = \sum_{q=1}^{Q} w^{(q)}(t) (c_1^{(q)}(t) \cos(\mu^{(q)^T}(t)t) \\
 + c_2^{(q)}(t) \sin(\mu^{(q)^T}(t)t)),
\]

\[
 c_1^{(q)}, c_2^{(q)} \sim \mathcal{GP}(0, k^{(Gibbs)}(t, t')).
\]
• SMK and extensions assume \textit{parametric} model.
• SMK and extensions assume *parametric* model.

• More flexible to use *nonparametric* model:

\[ s(\omega) = |\hat{h}(\omega)|^2. \]
• SMK and extensions assume \textit{parametric} model.

• More flexible to use \textit{nonparametric} model:

\[ s(\omega) = |\hat{h}(\omega)|^2. \]

• Inverse Fourier transform gives kernel:

\[ k(t, t') = \int_{-\infty}^{\infty} h(t - z)h(t' - z) \, dz = h \ast R(h)(t - t'). \]
• SMK and extensions assume parametric model.

• More flexible to use nonparametric model:

\[ s(\omega) = |\hat{h}(\omega)|^2. \]

• Inverse Fourier transform gives kernel:

\[ k(t, t') = \int_{-\infty}^{\infty} h(t - z)h(t' - z) \, dz = h \ast R(h)(t - t'). \]

• GPCM (Tobar et al., 2015) models \( h \sim \mathcal{GP}(0, k_h). \)
• SMK and extensions assume **parametric** model.

• More flexible to use **nonparametric** model:

  \[ s(\omega) = |\hat{h}(\omega)|^2. \]

• Inverse Fourier transform gives kernel:

  \[ k(t, t') = \int_{-\infty}^{\infty} h(t - z)h(t' - z) \, dz = h * R(h)(t - t'). \]

• GPCM (Tobar et al., 2015) models \( h \sim \mathcal{GP}(0, k_h) \).

  • \( \int_{-\infty}^{\infty} k_h(t, t) \, dt < \infty \) (finite trace).
• Nonparametric prior over kernels and PSDs.
• Interpretation as linear system:
• Interpretation as linear system:

\[
\begin{align*}
\text{white noise} & \overset{\rightarrow}{\rightarrow} h(t) & \overset{\rightarrow}{\rightarrow} f(t), \\
\text{white noise} & \sim \mathcal{GP}(0, \delta(t - t')), \\
\quad h & \sim \mathcal{GP}(0, k_h). 
\end{align*}
\]
• Interpretation as linear system:

![Diagram]

white noise \rightarrow h(t) \rightarrow f(t),

white noise \sim \mathcal{GP}(0, \delta(t - t')),

h \sim \mathcal{GP}(0, k_h).

• Inference complicated.
• Instead of designing kernel, design PSD.
• Instead of designing kernel, design PSD.

• Parametric approaches:
  • line spectrum (SSA),
  • mixture of Gaussians (SMK, MOSMK, GSMK).
• Instead of designing kernel, design PSD.

• Parametric approaches:
  • line spectrum (SSA),
  • mixture of Gaussians (SMK, MOSMK, GSMK).

• Nonparametric approach also possible (GPCM).